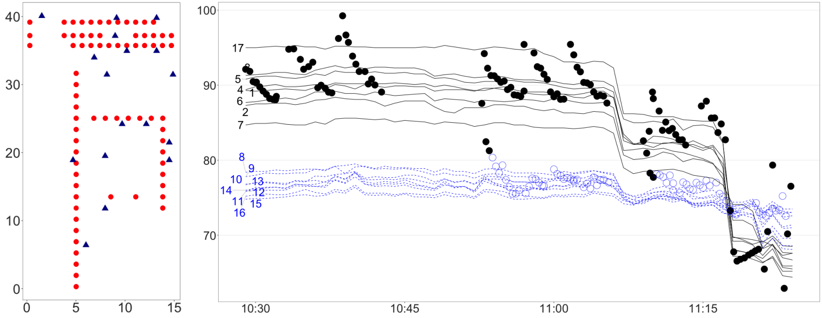
Semiparametric Spatio-temporal Modeling and Inference Under Local Stationarity

Jialuo Liu

Department of Statistics
Colorado State University

Motivating Example: Noise Intensity Data



- 17 static sensors and 2 roving sensors.
- collected between 10:29:00 am and 11:24:00 am.
- 56 observations, one per minute, for each static sensor.
- measurements from the roving sensors are irregularly spaced in time.

Spatio-temporal Modeling

Challenges

- Irregularity in space and in time
- Stationary vs nonstationary
- Separable vs nonseparable
- Lack of an appropriate asymptotic framework

Solution

- Model
  - ⋆ linear model
  - ⋆ nonlinear model
  - ⋆ semiparametric model
- Covariance function
  - ⋆ locally stationary covariance function
- Properties of estimates
  - ⋆ a spatio-temporal asymptotic framework

Spatio-temporal Model

$$y(\boldsymbol{s},t) = \mu(\boldsymbol{s},t) + \varepsilon_1(\boldsymbol{s},t) + \varepsilon_2(\boldsymbol{s},t),\ \boldsymbol{s} \in \mathcal{R} = [0,1]^d,\ t \in \mathcal{T} = [0,1],$$

- $\mu(\boldsymbol{s},t)$ is a fixed unknown spatio-temporal mean function.
- $\varepsilon_1(\boldsymbol{s},t)$ is a spatio-temporal Gaussian error process.
- $\varepsilon_2(\boldsymbol{s},t)$ follows i.i.d. Gaussian distribution with mean 0 and variance τ^2, independent of $\varepsilon_1(\boldsymbol{s},t)$.
- Denote the covariance function of $\varepsilon(\boldsymbol{s},t) = \varepsilon_1(\boldsymbol{s},t) + \varepsilon_2(\boldsymbol{s},t)$ as $\gamma((\boldsymbol{s},t),(\boldsymbol{s}',t'))$ for $(\boldsymbol{s},t),(\boldsymbol{s}',t') \in \mathcal{R} \times \mathcal{T}$.

Locally Stationary Covariance: General

(LS.1). There exists a sequence of functions $g_n(\cdot,\cdot,\boldsymbol{s},t)$ such that

$$|\gamma_n\left((\boldsymbol{s},t),(\boldsymbol{s}',t')\right) - g_n\left(\boldsymbol{s}'-\boldsymbol{s},t'-t,\boldsymbol{s},t\right)| = \mathcal{O}\left(\|\boldsymbol{s}'-\boldsymbol{s}\| + |t'-t| + \rho_n\right)$$

uniformly for all $(\boldsymbol{s},t),(\boldsymbol{s}',t') \in \mathcal{R}\times\mathcal{T}$, where $\{\rho_n\}$ is a sequence of positive numbers such that $\rho_n \to 0$ as $n\to\infty$. In addition, we assume that there exists a function g such that

$$\lim_{n\to\infty} |g_n\left(\boldsymbol{s}'-\boldsymbol{s},t'-t,\boldsymbol{s},t\right) - g(\boldsymbol{u}_1,u_2,\boldsymbol{s},t)| \to 0, \text{ as } n\to\infty,$$

where $\boldsymbol{u}_1 = A_n(\boldsymbol{s}'-\boldsymbol{s})$ and $u_2 = B_n(t'-t)$.

Locally Stationary Covariance: General

(LS.2). Define $g(\boldsymbol{s}, t) = g(\boldsymbol{0}, 0, \boldsymbol{s}, t)$ with $g(\boldsymbol{u}_1, u_2, \boldsymbol{s}, t)$ in (LS.1), and $g(\boldsymbol{s}, t)$ is such that $|g(\boldsymbol{s}, t) - g(\boldsymbol{s}', t')| \leq C_1 \|\boldsymbol{s} - \boldsymbol{s}'\| + C_2 |t - t'|$ for all $(\boldsymbol{s}, t), (\boldsymbol{s}', t') \in \mathcal{R} \times \mathcal{T}$, where $C_1, C_2 > 0$ are constants.

(LS.3). There exist two positive nonincreasing functions γ_0 and γ_1 satisfying $\int_0^\infty u^{d-1} \gamma_0(u) du < \infty$ and $\int_0^\infty \gamma_1(u) du < \infty$ such that $|\gamma_n((\boldsymbol{s}, t), (\boldsymbol{s} + \boldsymbol{u}_1/A_n, t + u_2/B_n))| \leq \gamma_0(\|\boldsymbol{u}_1\|)\gamma_1(|u_2|)$ for all n and $\|\boldsymbol{u}_1\|, |u_2| \in [0, \infty)$ such that $(\boldsymbol{s}, t), (\boldsymbol{s} + \boldsymbol{u}_1/A_n, t + u_2/B_n) \in \mathcal{R} \times \mathcal{T}$.

Some Notation of Derivatives

For a parametric covariance function determined by a $q \times 1$ vector $\boldsymbol{\theta}$, we denoted it as $\gamma_n(\cdot, \cdot; \boldsymbol{\theta})$.

- $\gamma_{n,k}(\cdot, \cdot; \boldsymbol{\theta}) = \partial \gamma_n(\cdot, \cdot; \boldsymbol{\theta}) / \partial \theta_k$, for $1 \leq k \leq q$.
- $\gamma_{n,kk'}(\cdot, \cdot; \boldsymbol{\theta}) = \partial \gamma_n(\cdot, \cdot; \boldsymbol{\theta}) / \partial \theta_k \partial \theta_{k'}$, for $1 \leq k, k' \leq q$.

If the data are observed at N_n sampling locations and time points $(\boldsymbol{s}_1, t_1), \ldots, (\boldsymbol{s}_{N_n}, t_{N_n})$. Let $\boldsymbol{\Gamma} = [\gamma_n((\boldsymbol{s}_i, t_i), (\boldsymbol{s}_j, t_j); \boldsymbol{\theta})]_{i,j=1}^{N_n}$ denote an $N_n \times N_n$ covariance matrix.

- $\boldsymbol{\Gamma}_k = \partial \boldsymbol{\Gamma} / \partial \theta_k$
- $\boldsymbol{\Gamma}_{kk'} = \partial^2 \boldsymbol{\Gamma} / \partial \theta_k \partial \theta_{k'}$

Locally Stationary Covariance: Parametric

(LS.4). The covariance function $\gamma_n(\cdot, \cdot; \boldsymbol{\theta})$ is bounded and is twice continuously differentiable with respect to $\boldsymbol{\theta}$ in an open set.

(LS.5). There exist two positive nonincreasing functions γ_2 and γ_3 with $\int_0^\infty u^{d-1}\gamma_2(u)du < \infty$ and $\int_0^\infty \gamma_3(u)du < \infty$ such that $\max\{|\gamma_{n,k}((\boldsymbol{s}, t), (\boldsymbol{s} + \boldsymbol{u}_1/A_n, t + u_2/B_n))|, |\gamma_{n,kk'}((\boldsymbol{s}, t), (\boldsymbol{s} + \boldsymbol{u}_1/A_n, t + u_2/B_n))|\} \leq \gamma_2(\|\boldsymbol{u}_1\|)\gamma_3(|u_2|)$ for all n and $\|\boldsymbol{u}_1\|, |u_2| \in [0, \infty)$ such that $(\boldsymbol{s}, t), (\boldsymbol{s} + \boldsymbol{u}_1/A_n, t + u_2/B_n) \in \mathcal{R} \times \mathcal{T}$ and $1 \leq k, k' \leq q$.

An Example

$$\gamma_n((\boldsymbol{s},t),(\boldsymbol{s}',t');\boldsymbol{\theta}) = \begin{cases} \frac{D(\boldsymbol{s},t)D(\boldsymbol{s}',t')\sigma^2\theta_3^{d/2}2^{1-\nu}}{(\theta_1^2u_2^2+1)^\nu(\theta_1^2u_2^2+\theta_3)^{d/2}\Gamma(\nu)} m(\boldsymbol{u}_1,u_2)^\nu K_\nu\left\{m(\boldsymbol{u}_1,u_2)\right\}, & \text{if } \|\boldsymbol{u}_1\|>0, \\ \frac{D(\boldsymbol{s},t)D(\boldsymbol{s}',t')\sigma^2\theta_3^{d/2}}{(\theta_1^2u_2^2+1)^\nu(\theta_1^2u_2^2+\theta_3)^{d/2}}, & \text{if } \|\boldsymbol{u}_1\|=0, |u_2|>0, \\ D(\boldsymbol{s},t)^2\sigma^2+\tau^2, & \text{if } \|\boldsymbol{u}_1\|=0, |u_2|=0, \end{cases}$$

- $\boldsymbol{u}_1 = \varrho_{1,n}(\boldsymbol{s}'-\boldsymbol{s})$, $u_2 = \varrho_{2,n}(t'-t)$,
- $m(\boldsymbol{u}_1,u_2) = \theta_2\left(\frac{\theta_1^2u_2^2+1}{\theta_1^2u_2^2+\theta_3}\right)^{1/2}\|\boldsymbol{u}_1\|$,
- $K_\nu(\cdot)$ is the modified Bessel function of the second kind of order ν,
- $D(\boldsymbol{s},t)$ is some fixed positive spatio-temporal function.

Proposition

Let $D(\boldsymbol{s},t)$ be some positive known function with $D(\boldsymbol{0},0)=1$ and $|D(\boldsymbol{s},t)-D(\boldsymbol{s}',t')| \le \widetilde{C}_1\|\boldsymbol{s}-\boldsymbol{s}'\| + \widetilde{C}_2|t-t'|$ for all $(\boldsymbol{s},t),(\boldsymbol{s}',t') \in \mathcal{R}\times\mathcal{T}$, where $\widetilde{C}_1,\widetilde{C}_2>0$ are constants. Then the generalized spatio-temporal Matérn covariance function (9) *satisfies conditions (LS.1)–(LS.5).*

Asymptotics in Spatial Statistics

Two traditional asymptotic frameworks

- Increasing domain asymptotic framework
  - ⋆ An increasing domain asymptotic framework does not permit the study of the "local" behavior of the covariance function.
- Infill domain asymptotic framework
  - ⋆ Some parameters in the covariance function are not consistently estimable under infill asymptotics.

(L_n, T_n)-rate STDE Asymptotics

Assume that, for all n,

(A.1). $\delta_n / \min_{1 \le j \le N_n} \delta_{j,n} \le c_1$,

(A.2). $\zeta_n / \min_{1 \le j \le N_n} \zeta_{j,n} \le c_2$,

(A.3). $\delta_n^d A_n^d \zeta_n B_n \ge c_3$,

- $\{A_n\}$ and $\{B_n\}$ are two sequences of positive numbers, both tend to infinity as n increases.
- $\delta_{j,n} = \min\{\|\boldsymbol{s}_i - \boldsymbol{s}_j\| : 1 \le i \le N_n, \boldsymbol{s}_i \ne \boldsymbol{s}_j\}$, $\delta_n = \max_{1 \le j \le N_n} \delta_{j,n}$.
- $\zeta_{j,n} = \min\{|t_i - t_j| : 1 \le i \le N_n, t_i \ne t_j\}$, $\zeta_n = \max_{1 \le j \le N_n} \zeta_{j,n}$.
- c_1, c_2, and c_3 are some positive constants independent of n.

We refer to (A.1)-(A.3) as an (A_n, B_n)*-rate spatio-temporal distance expanding asymptotics for fixed spatio-temporal domain* (STDE).

Commonly-used Spatio-temporal Models

- Zero-mean Gaussian process model: $\mu(\boldsymbol{s}, t) \equiv 0$.
- Simple linear regression model: $\mu(\boldsymbol{s}, t) = \boldsymbol{x}(\boldsymbol{s}, t)^\top \boldsymbol{\beta}$.
- Partially linear model: $\mu(\boldsymbol{s}, t) = \boldsymbol{x}(\boldsymbol{s}, t)^\top \boldsymbol{\beta} + f(t)$.

Zero-mean Gaussian Process Model

Model

$$y(\boldsymbol{s}, t) = \varepsilon_1(\boldsymbol{s}, t) + \varepsilon_2(\boldsymbol{s}, t), \quad \boldsymbol{s} \in \mathcal{R},\ t \in \mathcal{T},$$

Log-likelihood Function

$$\ell_{\mathrm{zm}}(\boldsymbol{\theta}) = -(N_n/2)\log(2\pi) - (1/2)\log\{\det \boldsymbol{\Gamma}(\boldsymbol{\theta})\} - (1/2)\boldsymbol{y}^\top \boldsymbol{\Gamma}(\boldsymbol{\theta})^{-1}\boldsymbol{y}.$$

The maximizer of $\ell_{\mathrm{zm}}(\boldsymbol{\theta})$, denoted by $\widehat{\boldsymbol{\theta}}_{\mathrm{MLE,zm}}$, is the maximum likelihood estimate of $\boldsymbol{\theta}$.

Consistency and Asymptotic Normality, MLE

Theorem

Under (LS.1)–(LS.5) and some regularity conditions, there exists, with probability tending to one, a local maximizer ${}^n\widehat{\boldsymbol{\theta}}_{\mathrm{zm}}$ *of* $\ell_{\mathrm{zm}}(\boldsymbol{\theta})$ *such that* $\|{}^n\widehat{\boldsymbol{\theta}}_{\mathrm{zm}} - \boldsymbol{\theta}_0\| = \mathcal{O}_p(N_n^{-1/2})$*. Moreover, the local maximizer* ${}^n\widehat{\boldsymbol{\theta}}_{\mathrm{zm}}$ *is asymptotic normal; as* $n \to \infty$*,*

$$N_n^{1/2}({}^n\widehat{\boldsymbol{\theta}}_{\mathrm{zm}} - \boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\mathcal{I}}_{\mathrm{zm}}(\boldsymbol{\theta}_0)^{-1}).$$

Model

$$y(\boldsymbol{s},t) = \boldsymbol{x}(\boldsymbol{s},t)^\top \boldsymbol{\beta} + \varepsilon_1(\boldsymbol{s},t) + \varepsilon_2(\boldsymbol{s},t), \quad \boldsymbol{s} \in \mathcal{R},\ t \in \mathcal{T},$$

- $\boldsymbol{x}(\boldsymbol{s},t) = (x_1(\boldsymbol{s},t), \ldots, x_p(\boldsymbol{s},t))^\top$,
- $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$.

Let $\boldsymbol{X} = [x_j(\boldsymbol{s}_i, t_i)]_{i=1,j=1}^{N_n,p}$ and $\boldsymbol{\eta} = (\boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top)^\top$,

Log-likelihood Function

$$\begin{aligned}\ell_{\text{reg}}(\boldsymbol{\eta}) = &-(N_n/2)\log(2\pi) - (1/2)\log\{\det \boldsymbol{\Gamma}(\boldsymbol{\theta})\} \\ &-(1/2)(\boldsymbol{y} - \boldsymbol{X}\beta)^\top \boldsymbol{\Gamma}(\boldsymbol{\theta})^{-1}(\boldsymbol{y} - \boldsymbol{X}\beta).\end{aligned}$$

Denote the maximizer of $\ell_{\text{reg}}(\boldsymbol{\eta})$ as $\widehat{\boldsymbol{\eta}}_{\text{MLE,reg}} = (\widehat{\boldsymbol{\beta}}_{\text{MLE,reg}}, \widehat{\boldsymbol{\theta}}_{\text{MLE,reg}})$.

Consistency and Asymptotic Normality, MLE

Theorem

Under (LS.1)–(LS.5) and some regularity conditions, there exists, with probability tending to one, a local maximizer ${}^{n}\widehat{\boldsymbol{\eta}}_{\mathrm{reg}} = ({}^{n}\widehat{\boldsymbol{\beta}}_{\mathrm{reg}}{}^{\top}, {}^{n}\widehat{\boldsymbol{\theta}}_{\mathrm{reg}}{}^{\top})^{\top}$ *of* $\ell_{\mathrm{reg}}(\boldsymbol{\eta})$ *such that* $\|{}^{n}\widehat{\boldsymbol{\eta}}_{\mathrm{reg}} - \boldsymbol{\eta}_0\| = \mathcal{O}_p(N_n^{-1/2})$*. Moreover, the local maximizer* ${}^{n}\widehat{\boldsymbol{\eta}}_{\mathrm{reg}}$ *is asymptotic normal; as* $n \to \infty$*,*

$$N_n^{1/2}({}^{n}\widehat{\boldsymbol{\beta}}_{\mathrm{reg}} - \boldsymbol{\beta}_0) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\mathcal{I}}_{\mathrm{reg}}(\boldsymbol{\beta}_0)^{-1}) \quad \textit{and}$$
$$N_n^{1/2}({}^{n}\widehat{\boldsymbol{\theta}}_{\mathrm{reg}} - \boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\mathcal{I}}_{\mathrm{reg}}(\boldsymbol{\theta}_0)^{-1}).$$

Partially Linear Model

Model

$$y(\boldsymbol{s},t) = \boldsymbol{x}(\boldsymbol{s},t)^\top \boldsymbol{\beta} + f(t) + \varepsilon_1(\boldsymbol{s},t) + \varepsilon_2(\boldsymbol{s},t), \quad \boldsymbol{s} \in \mathcal{R},\ t \in \mathcal{T},$$

- $f(t) = 0$: simple linear regression model, use MLE.
- $f(t) \neq 0$: MLE cannot be used directly.

Profile Likelihood Estimation

- Let $y_i^* = y(\boldsymbol{s}_i, t_i) - \boldsymbol{x}(\boldsymbol{s}_i, t_i)^\top \boldsymbol{\beta}$ and $\boldsymbol{y}^* = (y_1^*, \ldots, y_{N_n}^*)$.
- Using local polynomial regression, minimizing

$$\sum_{i=1}^{N_n} \{y_i^* - b_{0,t} - b_{1,t}(t_i - t)\}^2 K_h(t_i - t),$$

  with respect to $\boldsymbol{b}_t = (b_{0,t}, b_{1,t})^\top$, where h is a bandwidth, $K_h = K(\cdot/h)/h$ with a kernel function $K(\cdot)$.
- $(\widehat{b}_{0,t}, h\widehat{b}_{1,t})^\top = \boldsymbol{\omega}(t)\boldsymbol{y}^*$, where $\boldsymbol{\omega}(t) = (\boldsymbol{D}_t^\top \boldsymbol{K}_t \boldsymbol{D}_t)^{-1} \boldsymbol{D}_t^\top \boldsymbol{K}_t$.
  - $\boldsymbol{K}_t = \mathrm{diag}(K_h(t_1 - t), \ldots, K_h(t_{N_n} - t))$,
  - $\boldsymbol{D}_t = (\mathbf{1}_{N_n}, \boldsymbol{d}_{1t})$, where $\mathbf{1}_{N_n}$ is an $N_n \times 1$ vector of 1's, and $\boldsymbol{d}_{1t} = \left(\frac{t_1 - t}{h}, \ldots, \frac{t_{N_n} - t}{h}\right)^\top$.

Profile Likelihood Estimation

- Let $\boldsymbol{f} = (f(t_1), \ldots, f(t_{N_n}))^\top$ and $\boldsymbol{f}' = (f'(t_1), \ldots, f'(t_{N_n}))^\top$.
- $\widetilde{\boldsymbol{f}} = \boldsymbol{S}\boldsymbol{y}^* = \boldsymbol{S}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$, where the smoother matrix is

$$\boldsymbol{S} = \left(\boldsymbol{\omega}_1(t_1)^\top, \ldots, \boldsymbol{\omega}_1(t_{N_n})^\top\right)^\top,$$

  where $\boldsymbol{\omega}_1(t) = (1, 0)\boldsymbol{\omega}(t)$.
- Plugging $\widetilde{\boldsymbol{f}}$, $(\boldsymbol{I} - \boldsymbol{S})\boldsymbol{y} \approx (\boldsymbol{I} - \boldsymbol{S})\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$,
- The estimator $(\widehat{\boldsymbol{\beta}}^\top, \widehat{\boldsymbol{\theta}}^\top)^\top$ is obtained by maximizing

$$-\frac{1}{2}\log\{\det \boldsymbol{\Gamma}(\boldsymbol{\theta})\} - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{I} - \boldsymbol{S})^\top\boldsymbol{\Gamma}(\boldsymbol{\theta})^{-1}(\boldsymbol{I} - \boldsymbol{S})(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}),$$

- $\widehat{\boldsymbol{f}} = \boldsymbol{S}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})$

Consistency and Asymptotic Normality, PMLE (1)

Theorem

Under (LS.1)–(LS.5) and some regularity conditions, there exists, with probability tending to one, a local maximizer ${}^n\widehat{\boldsymbol{\eta}} = ({}^n\widehat{\boldsymbol{\beta}}^\top, {}^n\widehat{\boldsymbol{\theta}}^\top)^\top$ *of* $\ell(\boldsymbol{\eta})$ *such that* $\|{}^n\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0\| = \mathcal{O}_p(N_n^{-1/2})$. *Moreover, the local maximizer* ${}^n\widehat{\boldsymbol{\eta}}$ *is asymptotic normal; as* $n \to \infty$,

$$N_n^{1/2}({}^n\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Pi}^{-1}) \quad \textit{and}$$

$$N_n^{1/2}({}^n\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\mathcal{I}}_0(\boldsymbol{\theta}_0)^{-1}).$$

- The estimate ${}^n\widehat{\boldsymbol{\eta}}$ is root-N_n consistent.
- The asymptotic variance of $\widehat{\boldsymbol{\beta}}$ is not $\boldsymbol{\mathcal{I}}_0(\boldsymbol{\beta}_0)^{-1}$, i.e., the asymptotic variance as if $\boldsymbol{f}$ is given.
- For ${}^n\widehat{\boldsymbol{\theta}}$, we have the same asymptotic variance as that of the case when $\boldsymbol{f}$ is given.

Consistency and Asymptotic Normality, PMLE (2)

(C.1). Given $t \in (0,1)$, there exists a 2×2 matrix $\boldsymbol{\Delta}_t$, such that $(N_n^{-1}h)\boldsymbol{k}_t^\top \boldsymbol{\Gamma} \boldsymbol{k}_t \longrightarrow q(t)^2 \boldsymbol{\Delta}_t$, where $\boldsymbol{k}_t = \{K_h(t_i - t)\{(t_i - t)/h\}^{j-1}\}_{i,j=1}^{N_n,2}$ is an $N_n \times 2$ matrix.

Theorem

If $f^{(3)}(t)$ is bounded, under (LS.1)–(LS.5), (C.1) and some regularity assumptions, we have, as $n \to \infty$,

$$(N_n h)^{1/2} \left\{ \widehat{\boldsymbol{F}}(t) - \boldsymbol{F}(t) - (1/2)h^2 \begin{pmatrix} \mu_2 f''(t) \\ 0 \end{pmatrix} + o(h^2) \right\} \xrightarrow{D} N\left(\mathbf{0}, \begin{pmatrix} 1 & 0 \\ 0 & \mu_2^{-1} \end{pmatrix} \boldsymbol{\Delta}_t \begin{pmatrix} 1 & 0 \\ 0 & \mu_2^{-1} \end{pmatrix}\right),$$

for $t \in (0,1)$, where $\mu_k = \int_{-\infty}^{\infty} x^k K(x) dx$.

Optimal Bandwidth

- The asymptotic mean squared error (AMSE) of $\widehat{f}(t)$ is

$$\mathrm{AMSE}(t) = (1/4)h^4\mu_2^2 f''(t)^2 + (N_n h)^{-1}(1,0)\boldsymbol{\Delta}_t(1,0)^\top.$$

- The asymptotic weighted mean integrated squared error is

$$\begin{aligned}\mathrm{AMISE}(h) = \int_0^1 \mathrm{AMSE}(t)q(t)dt = (1/4)h^4\mu_2^2 \int_0^1 f''(t)^2 q(t)dt \\ + (N_n h)^{-1}\int_0^1 (1,0)\boldsymbol{\Delta}_t(1,0)^\top q(t)dt.\end{aligned}$$

- We obtain an asymptotically optimal bandwidth as

$$h_{\mathrm{opt}} = N_n^{-1/5}\mu_2^{-2/5}\left\{\frac{\int_0^1 (1,0)\boldsymbol{\Delta}_t(1,0)^\top q(t)dt}{\int_0^1 f''(t)^2 q(t)dt}\right\}^{1/5}.$$

- The convergence rate is $N_n^{2/5}$ and is the nonparametric optimal rate.

Some Problems of Optimal Bandwidth

- The asymptotically optimal bandwidth h_{opt} depends on several unknown quantities:
  - $\boldsymbol{\Delta}_t$ in the asymptotic variance of $\widehat{\boldsymbol{F}}(t)$,
  - the density of sampling time points $q(t)$, and
  - the second-order derivative of the temporal function $f''(t)$.
- It is not practical to estimate h_{opt}, and a more practical bandwidth selection procedure is needed.
- We propose a bandwidth selection procedure based on cross-validation.

Choice of Kernels and Bandwidth Selection

- Cross-validation method has been widely used for independent and identically distributed data.
- For correlated data, (De Brabanter et al. [2011]) showed that bandwidths selected by cross-validation methods with commonly-used kernels are biased, and tend to choose a much smaller bandwidth than the optimal one when data are positively correlated.
- Bimodal kernels are used to address the bias in bandwidth selection, and theoretical properties are established for one-dimensional grid data with stationary covariance (e.g. AR(1)).

Theorem

Under some regularity conditions, for the spatio-temporal partially linear model with a covariance function satisfying (LS.1)–(LS.5), if there exists a sequence $C_n > 0$ *such that* $C_n h^{-1} \to 0$ *and* $1/(B_n \zeta_n) \int_{B_n C_n}^{\infty} \gamma_1(u) du \to 0$ *as* $n \to \infty$*, then we have*

$$\begin{aligned} \mathrm{E}\{\mathrm{CV}(h)\} =& N_n^{-1} \sum_{i=1}^{N_n} \mathrm{E}\{f(t_i) - \widehat{f}^{(-i)}(t_i)\}^2 + \overline{\sigma^2} \\ & - K(0) \left(\frac{2}{N_n} \sum_{i=1}^{N_n} \sum_{\substack{j \neq i \\ |t_j - t_i| < C_n}} \frac{\mathrm{Cov}(\varepsilon_i, \varepsilon_j)}{b(t_i) - K(0)} \right) + o\left(\frac{1}{N_n h}\right), \end{aligned}$$

where $\mathrm{CV}(h) = N_n^{-1} \sum_{i=1}^{N_n} \left\{ y_i^* - \widehat{f}^{(-i)}(t_i) \right\}^2$, $\overline{\sigma^2} = N_n^{-1} \sum_{i=1}^{N_n} \mathrm{Var}(Y_i)$ *and* $b(t_i) = N_n q(t_i) h (\mu_{0,t_i} \mu_{2,t_i} - \mu_{1,t_i}^2) \mu_{2,t_i}^{-1}$.

Bimodal Kernel

We consider the bimodal kernel [De Brabanter et al., 2011]:

$$K_2(u) = \frac{2}{\sqrt{\pi}} u^2 \exp(-u^2).$$

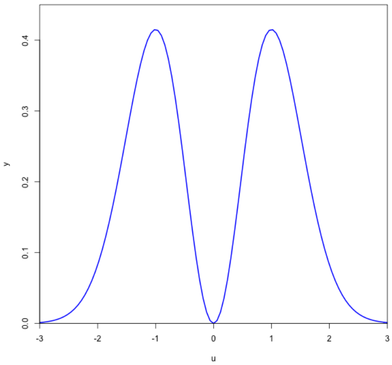


Figure 2: K_2 kernel

Bandwidth Selection Procedure

We propose the following procedure for the selection of bandwidth h.

(1) For a predetermined bandwidth h_0 and a kernel function K, obtain the estimated regression coefficients $\widetilde{\boldsymbol{\beta}}$ by the profile likelihood method.

(2) For given a kernel function, find the bandwidth h_{opt} that minimizes the cross-validation criterion

$$\mathrm{CV}(h) = N_n^{-1} \sum_{i=1}^{N_n} \left(\frac{\widetilde{y}_i^* - \widetilde{f}_i^*}{1 - S_{ii}} \right)^2, \quad (1)$$

where

- $\widetilde{y}_i^* = y_i - \boldsymbol{x}(\boldsymbol{s}_i, t_i)^\top \widetilde{\boldsymbol{\beta}}$, and
- $\widetilde{f}_i^*$ is the profile likelihood estimate.

(3) Use h_{opt} and the kernel function from Step (2) to obtain the desired estimates of both the regression coefficients $\boldsymbol{\beta}$ and the covariance function parameters $\boldsymbol{\theta}$.

Other Bandwidth Selection Criteria

- A popular alternative to the cross-validation criterion (1) is the generalized cross-validation [GCV; Golub et al., 1979] criterion, in which S_{ii} is replaced by $N_n^{-1}\,\mathrm{tr}(S)$.
- For dependent data, Francisco-Fernandez and Opsomer [2005] proposed a bias-corrected generalized cross-validation criterion (GCV_c), replacing S_{ii} by $N_n^{-1}\,\mathrm{tr}(SR(\boldsymbol{\theta}))$; that is,

$$\mathrm{GCV}_{\mathrm{c}}(h) = N_n^{-1} \sum_{i=1}^{N_n} \left(\frac{\widetilde{y}_i^* - \widetilde{f}_i^*}{1 - N_n^{-1}\,\mathrm{tr}(SR(\boldsymbol{\theta}))} \right)^2 , \qquad (2)$$

- In practice, a pilot estimate of the covariance parameters is required.

Simulations Set-up

- N_s sampling locations, $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{N_s}$, within the spatial domain $[0,1]^2$.
- At each sampling location, we consider N_t time points $t_1, \ldots, t_{N_t}$, where $t_i = (i - 1/2)/N_t$ for $i = 1, \ldots, N_t$ and $N_t = 1000$, and each time point has a probability 0.04 being sampled.
- The sampling locations and the sampling time points are generated once and remain fixed throughout the simulation study.
- $\boldsymbol{\beta} = (4, 3, 2, 1)^\top$.
- The covariates are drawn (once) from a multivariate normal distribution with zero mean, unit variance, and a cross-covariate correlation of 0.5.
- $f(t) = 2(1 - \cos(2\pi t))$.
- 400 iterations.

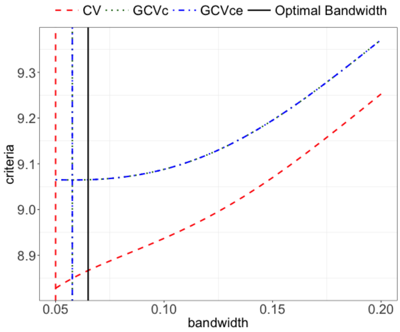
Simulation Set-up

The first covariance structure: an exponential spatio-temporal covariance function

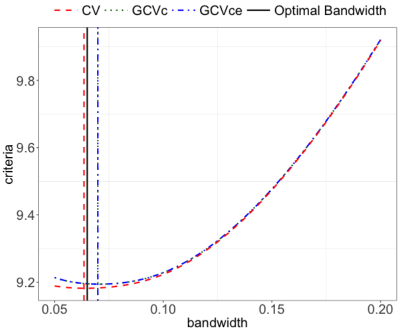
$$\mathrm{Cov}\{\epsilon(\boldsymbol{s}_i, t_i), \epsilon(\boldsymbol{s}_j, t_j)\} = \begin{cases} \sigma^2(1-c)\exp\{-\varrho_{1,n}\|\boldsymbol{s}_i - \boldsymbol{s}_j\|/c_s - \varrho_{2,n}|t_i - t_j|/c_t\}, & \text{if } i \neq j; \\ \sigma^2, & \text{if } i = j. \end{cases}$$

- σ^2 is the variance of the error process,
- $c \in [0, 1]$ is a nugget proportion such that $c\sigma^2$ is the nugget effect,
- c_s and c_t are positive range parameters in space and time, respectively.
- $\sigma^2 = 9.0, c = 0.2, c_s = 1$ and $c_t = 1$.

Simulation Results: Bandwidth Selection



(a) Gaussian Kernel.



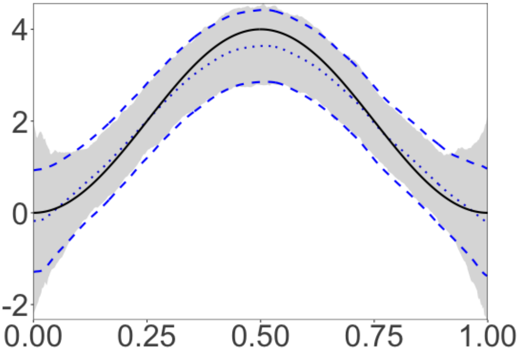
(b) Bimodal Kernel

Simulation Results: Parameter Estimates

Table 1: Sample mean, sample standard deviation (`SD`), averaged estimated standard deviation (`SDm`) of regression and covariance parameters, and mean-squared prediction error (MSPE) for $N_s = 20$ for `COV-1`.

Method	Truth	PLM			ALT_1			ALT_2		
		mean	SD	SDm	mean	SD	SDm	mean	SD	SDm
β_1	4.0	3.985	0.104	0.112	3.991	0.120	0.132	3.985	0.104	0.112
β_2	3.0	3.017	0.123	0.116	3.017	0.140	0.136	3.018	0.122	0.117
β_3	2.0	2.006	0.109	0.114	2.003	0.127	0.132	2.005	0.109	0.114
β_4	1.0	0.996	0.110	0.115	0.988	0.131	0.134	0.996	0.110	0.115
σ^2	9.0	9.111	0.569	0.546	9.101	0.589	0.453	8.944	0.525	0.528
c	0.2	0.209	0.077	0.074	—	—	—	0.202	0.078	0.077
c_s	1.0	1.090	0.224	0.213	—	—	—	1.025	0.198	0.200
c_t	1.0	1.087	0.243	0.213	—	—	—	1.029	0.224	0.204
MSPE	—	6.501			9.245			6.484		

Simulation Results: Nonparametric Estimation



- solid line: true temporal function $f(t) = 2(1 - \cos(2\pi t))$.
- dotted line: average of estimated temporal function
- dashed line: 95% pointwise estimated confidence intervals.
- grey region: 95% pointwise empirical confidence intervals

Simulation Results: Nonparametric Estimation

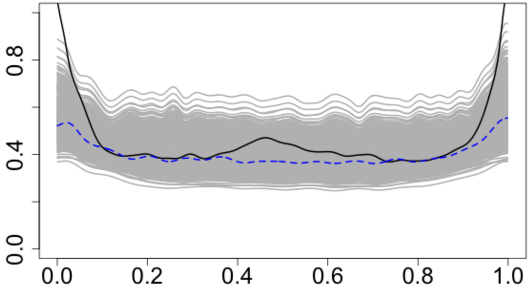


Figure 4: Estimated standard deviations

- solid line: pointwise simulation standard deviation
- dashed line: pointwise estimated standard deviation using true covariance parameters
- grey lines: pointwise estimated standard deviation from each iteration

Other Covariance Functions

Consider the generalized spatio-temporal Matérn covariance function

$$\mathrm{Cov}\{\epsilon(\boldsymbol{s}_i, t_i), \epsilon(\boldsymbol{s}_j, t_j)\} = \begin{cases} D(\boldsymbol{s}_i, t_i) D(\boldsymbol{s}_j, t_j) \frac{\sigma^2}{(a^2|T_n(t_i - t_j)|^2 + 1)^{3/2}} \exp\{-bL_n \|\boldsymbol{s}_i - \boldsymbol{s}_j\|\}, & \text{if } i \neq j; \\ D(\boldsymbol{s}_i, t_i) D(\boldsymbol{s}_j, t_j) \sigma^2 + c\sigma^2, & \text{if } i = j. \end{cases}$$

- σ^2 is the variance of the error process,
- $c \in [0, 1]$ is a nugget proportion such that $c\sigma^2$ is the nugget effect,
- a and b are positive range parameters in time and space, respectively,
- `COV-2`:
  - ⋆ $D(\boldsymbol{s}_i, t_i) = dt_i + 1$
  - ⋆ $\sigma^2 = 9, c = 0.2, a = 1, b = 1, d = 1$
- `COV-3:`
  - ⋆ $D(\boldsymbol{s}_i, t_i) = dt_i + es_{1i} + fs_{2i} + 1.$
  - ⋆ $\sigma^2 = 9.0, c = 0.2, a = 1, b = 1, d = 0.5, e = 0.5, f = 0.5$

Simulation Results: Parameter Estimates, COV-2

Table 2: Sample mean, sample standard deviation (SD), averaged estimated standard deviation (SDm) of regression and covariance parameters, and mean-squared prediction error (MSPE) for $N_s = 20$ for COV-2.

Method	Truth	PLM			ALT_1			ALT_2		
		mean	SD	SDm	mean	SD	SDm	mean	SD	SDm
β_1	4.0	3.969	0.136	0.146	3.987	0.197	0.209	3.969	0.137	0.146
β_2	3.0	3.020	0.165	0.153	3.027	0.218	0.215	3.023	0.164	0.153
β_3	2.0	2.011	0.146	0.152	1.998	0.201	0.209	2.008	0.146	0.152
β_4	1.0	1.003	0.144	0.148	0.983	0.208	0.213	1.004	0.144	0.148
σ^2	9.0	9.075	1.610	1.569	22.751	1.658	1.133	8.934	1.543	1.544
c	0.2	0.229	0.121	0.102	—	—	—	0.227	0.120	0.103
a	1.0	0.980	0.117	0.101	—	—	—	0.996	0.118	0.104
b	1.0	0.973	0.139	0.131	—	—	—	1.001	0.140	0.135
d	1.0	1.020	0.246	0.238	—	—	—	1.021	0.242	0.237
MSPE	—	13.454			24.114			13.429		

Simulation Results: Parameter Estimates, COV-3

Table 3: Sample mean, sample standard deviation (SD), averaged estimated standard deviation (SDm) of regression and covariance parameters, and mean-squared prediction error (MSPE) for $N_s = 20$ for COV-3.

Method	Truth	PLM			ALT_1			ALT_2		
		mean	SD	SDm	mean	SD	SDm	mean	SD	SDm
β_1	4.0	3.964	0.154	0.165	3.984	0.220	0.238	3.963	0.154	0.165
β_2	3.0	3.025	0.184	0.171	3.029	0.248	0.246	3.027	0.183	0.172
β_3	2.0	2.014	0.162	0.172	1.994	0.230	0.238	2.011	0.161	0.172
β_4	1.0	1.002	0.163	0.168	0.981	0.238	0.242	1.003	0.163	0.168
σ^2	9.0	9.322	2.418	2.397	29.520	2.115	1.471	9.066	2.374	2.343
c	0.2	0.238	0.151	0.134	—	—	—	0.240	0.155	0.138
a	1.0	0.982	0.115	0.098	—	—	—	0.996	0.116	0.101
b	1.0	0.977	0.135	0.127	—	—	—	1.001	0.135	0.130
d	0.5	0.509	0.222	0.222	—	—	—	0.513	0.221	0.224
e	0.5	0.502	0.223	0.217	—	—	—	0.515	0.228	0.220
f	0.5	0.514	0.193	0.197	—	—	—	0.524	0.198	0.200
MSPE	—	13.107			23.673			13.062		

Nonseparable but Stationary Covariance Functions

Consider the generalized spatio-temporal Matérn covariance function

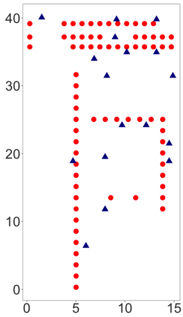
$$
\begin{aligned}
&\mathrm{Cov}\{\epsilon(\boldsymbol{s}_i, t_i), \epsilon(\boldsymbol{s}_j, t_j)\} \\
&= \begin{cases} \dfrac{D(\boldsymbol{s}_i,t_i)D(\boldsymbol{s}_j,t_j)\sigma^2(1-c)\theta_3 \exp\{-\theta_2 \left(\frac{\theta_1^2|\varrho_{2,n}(t_i-t_j)|^2+1}{\theta_1^2|\varrho_{2,n}(t_i-t_j)|^2+\theta_3}\right)^{1/2} \varrho_{1,n}\|\boldsymbol{s}_i-\boldsymbol{s}_j\|\}}{(\theta_1^2|\varrho_{2,n}(t_i-t_j)|^2+1)^{1/2}(\theta_1^2|\varrho_{2,n}(t_i-t_j)|^2+\theta_3)}, & \text{if } i \neq j; \\ D(\boldsymbol{s}_i, t_i)D(\boldsymbol{s}_j, t_j)\sigma^2 + c\sigma^2, & \text{if } i = j. \end{cases}
\end{aligned}
$$

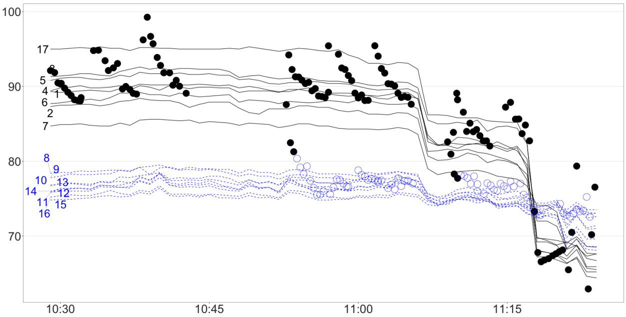
- σ^2 is the variance of the error process,
- $c \in [0, 1]$ is a nugget proportion such that $c\sigma^2$ is the nugget effect,
- θ_1 and θ_2 are positive range parameters in time and space, respectively,
- $\sigma^2 = 9, c = 0.2, \theta_1 = 1, \theta_2 = 1, \theta_3 = 4$.

Table 4: Sample mean, sample standard deviation (SD), averaged estimated standard deviation (SDm) of regression and covariance parameters, and mean-squared prediction error (MSPE) for $N_s = 20$.

Method	Truth	PLM			ALT_1			ALT_2		
		mean	SD	SDm	mean	SD	SDm	mean	SD	SDm
β_1	4.0	3.985	0.090	0.097	3.988	0.131	0.145	3.984	0.090	0.097
β_2	3.0	3.017	0.109	0.101	3.016	0.155	0.149	3.018	0.109	0.101
β_3	2.0	2.004	0.095	0.099	1.999	0.143	0.145	2.003	0.095	0.099
β_4	1.0	0.998	0.094	0.099	0.989	0.148	0.147	0.998	0.094	0.099
σ^2	9.0	9.235	0.949	0.890	10.955	2.995	—	8.990	0.871	0.857
c	0.2	0.198	0.045	0.043	—	—	—	0.201	0.046	0.045
θ_1	1.0	0.925	0.230	0.217	—	—	—	0.967	0.208	0.214
θ_2	1.0	0.962	0.421	0.441	—	—	—	1.039	0.415	0.473
θ_3	4.0	4.525	3.112	3.816	—	—	—	4.827	3.093	4.109
MSPE	—	4.915			11.206			4.908		

Real Data Analysis: Noise Intensity Data





Real Data Analysis

Model

$$y(\boldsymbol{s}, t) = \beta_1 s_1 + \beta_2 s_2 + f(t) + \varepsilon(\boldsymbol{s}, t), \quad \text{with } \boldsymbol{s} = (s_1, s_2).$$

Three Covariance Structures

- Stationary: $D_1(\boldsymbol{s}, t) = 1$
- Nonstationary:
  - $D_2(\boldsymbol{s}, t) = 1 + dt$.
  - $D_3(\boldsymbol{s}, t) = 1 + dt + e(t - \tau)_+$, τ is chosen as 11:02:00.

Real Data Analysis: Summary

	$D_1(\boldsymbol{s},t)$	$D_2(\boldsymbol{s},t)$	$D_3(\boldsymbol{s},t)$	$D_3(\boldsymbol{s},t)$ (penalized)
h	0.0193	0.0193	0.0193	0.0193
		Regression parameters		
β_1	-0.3922(0.0820)	-0.4492(0.0652)	-0.4569(0.0546)	-0.4600(0.0636)
β_2	0.3015(0.0565)	0.4048(0.0440)	0.4135(0.0413)	0.3954(0.0425)
		Covariance parameters		
σ^2	50.8840(8.7442)	8.8096(1.7254)	15.3138(2.9336)	14.7628(2.7797)
c	0.0007(0.0001)	0.0020(0.0005)	0.0010(0.0002)	0.0009(0.0002)
c_s	0.1662(0.0040)	0.1677(0.0037)	0.1654(0.0036)	0.1723(0.0038)
c_t	0.0152(0.0025)	0.0215(0.0033)	0.0191(0.0030)	0.0237(0.0036)
d	—	1.9218(0.2647)	0.3114(0.1514)	0.1982(0.1723)
e	—	—	12.1972(1.1460)	3.0394(0.4177)

- the estimate of the coefficient of $(t-\tau)_+$ in D_3 (e) is unusually large.
- a common phenomenon in spline smoothing with truncated polynomial basis functions.
- consider a penalized approach by adding an additional penalty term $-\lambda|e|$ to the log-likelihood function, where $\lambda = 20$.

Real Data Analysis

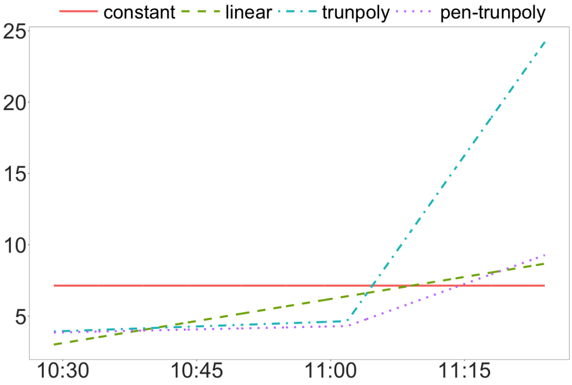
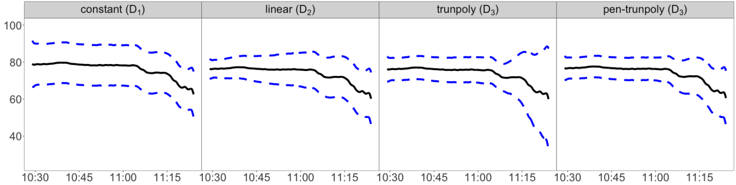


Figure 6: Estimated standard deviation

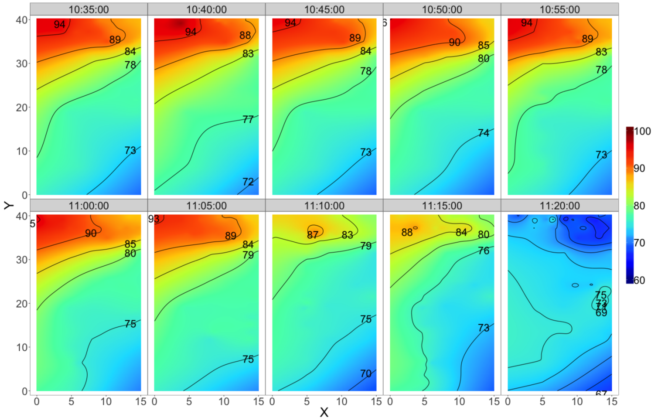
Real Data Analysis: Estimated Temporal Trend



- solid line: estimated temporal function
- dashed line: 95% pointwise estimated confidence intervals

- The function estimates $\widehat{f}(t)$ from D_1, D_2 and D_3 are similar.
- Penalized D_3 has the narrowest confidence intervals.

Kriging Map Based on D_3 (penalized)



- A possible noise source on the upper-left corner.
- A change of the overall noise intensity at 11:10:00.
- A horizontal separation around $y = 30$ before 11:10:00.

Reference

K. De Brabanter, J. De Brabanter, J. A. K. Suykens, and B. De Moor. Kernel regression in the presence of correlated errors. *Journal of Machine Learning Research*, 12:1955–1976, 2011.

M. Francisco-Fernandez and J. D. Opsomer. Smoothing parameter selection methods for nonparametric regression with spatially correlated errors. *Canadian Journal of Statistics*, 33: 279–295, 2005.

G. H. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223, 1979.